LightMover: Generative Light Movement with Color and Intensity Controls

Supplementary Material

A. Implementation Details

Training. The model is trained on 16 NVIDIA A100 GPUs for 15,000 iterations with a batch size of 32. We use the AdamW optimizer [1] with a weight decay of 0.01 and an initial learning rate of 1×10^{-4}. Exponential Moving Average (EMA) with a decay rate of 0.99 is applied after the first 1,000 iterations to stabilize training.

Data Sampling. Training samples are drawn at mixed resolutions of 512×512 and 1024×1024 pixels with a 1:1 ratio. The dataset combines synthetic and real data at a 10:1 ratio. Synthetic tasks are distributed across seven categories: (1) light movement, (2) object movement, (3) light color change, (4) light intensity change, (5) joint movement and color/intensity change, (6) light removal, and (7) light insertion, following a ratio of 6:3:3:3:1:1, respectively.

Augmentation. To improve robustness to spatial variation, we apply box augmentation by maintaining the bounding-box center while randomly scaling its dimensions within $[0.8, 1.2]$. We additionally apply light-illumination augmentation via the Physically Disentangled Rendering procedure (Sec. 3.2 of the main paper), which dynamically varies ambient and direct light components during training.

B. Additional Comparisons and Ablations

We present additional comparisons and ablation visualizations that complement the quantitative results in the main paper.

B.1. Comparison with Ground Truth

To validate the physical plausibility of LightMover, we directly compare our predictions against real ground-truth photographs from the *LightMove-A* benchmark, where each ground truth is captured by the same camera under identical ambient conditions with the light source physically relocated. As shown in Figure 1, LightMover achieves precise positional control for local lighting changes, accurately reproducing shadows, highlights, and surface shading patterns that closely match the ground-truth images. Notably, these examples are selected without cherry-picking from challenging in-the-wild scenes, demonstrating that our model generalizes robustly to diverse real-world environments without introducing visual hallucinations or physically inconsistent artifacts.

B.2. Visual Ablation of Training Data Composition

Figure 2 provides qualitative ablation results corresponding to the quantitative analysis in Table 4 of the main paper. These visual comparisons illustrate the effectiveness of the proposed data augmentation strategy and the synergistic effects of training data composition. Specifically, training with the physically disentangled rendering augmentation (*Light Aug.*) enables the model to generalize to complex shadow interactions, such as intricate leaf shading patterns, that are absent in the unaugmented data. Furthermore, incorporating light color and intensity variation tasks during training improves the model's understanding of how individual light sources contribute to the overall scene illumination. The full model, trained with all proposed data components, produces the most physically coherent results by effectively disentangling and recombining illumination effects.

B.3. Comparison with the Learnable Embedding Variant

We compare our frame-based conditioning design against a variant that uses separate learnable embeddings for different control signals (Table 5, line 3 in the main paper). As shown in Figure 3, we evaluate the joint move-with-color-change task, where the colored dot in the top-left corner of each result indicates the target light color. The variant with condition-specific learnable embeddings fails to correctly recognize color conditions when movement and color changes are applied simultaneously. This is because learnable embeddings are less flexible and do not generalize well to multiple simultaneous controls: they tend to conflate or ignore individual signals when multiple conditions are introduced jointly. In contrast, our frame-based conditioning naturally integrates with the video diffusion backbone and supports joint control over movement, color, and intensity by encoding each condition as an explicit image frame, enabling the model to maintain distinct and accurate representations for each control signal.

C. Prompt Details for Nano-Banana Light-Movement Baselines

As described in the main paper (Sec. 4), we evaluate two representative LLM-powered text-to-image editing setups for Gemini-2.5-Flash-Image: (1) a *one-step* editing pipeline, and (2) a *two-step* remove–insert pipeline. For completeness, we provide the exact prompt templates used in both settings, following the Gemini-2.5-Flash-Image's generation results.

C.1. One-Step Editing

In the one-step setup, both the original light-source location and the target location are marked in the input image using a red bounding box and a blue bounding box, respectively. The model receives a single prompt instructing it to relocate

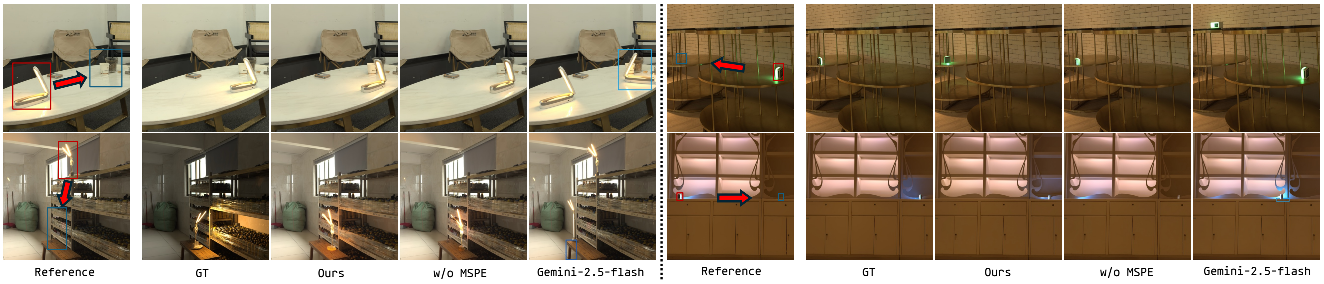

Figure 1. **Comparison with ground-truth images.** We show side-by-side comparisons between LightMover predictions and ground-truth photographs captured by the same camera under the same ambient conditions but with the light source physically relocated.

Figure 2. **Qualitative results for Table 4 (training data ablation).** With the proposed synthetic rendering data pipeline (Light Aug.), the model generalizes to complex shadow handling (e.g., leaf shading). Incorporating color and intensity variations further improves the modeling of light contributions and their combinations.

the light source from the red region to the blue region and produce the corresponding physically consistent relighting. We include three variants: *concise*, *basic*, and *precise* in our preliminary experiments. Detail prompts are shown below:

We show the generated results in Figure 7.

C.2. Two-Step Editing

For the two-step setup described in the main paper, we decompose the task into: (1) *light removal*, followed by (2) *light insertion at the target location*.

We first segment the light-source object using SAM and crop it into a standalone patch. The model is then prompted twice: once to remove the light source from the scene, and once to reinsert it at the blue target region with physically consistent relighting. We show the prompt for removal and insertion below.

We show the removal result and final editing results in Figure 8.

D. More Qualitative Results

To further analyze the behavior of our model under different illumination conditions, we present additional qualitative examples that highlight three key challenges in light movement: *surface interaction*, *cast shadows*, and *reflections*. These examples illustrate how our method jointly reasons about geometry, materials, and global illumination when the light source is repositioned.

D.1. Surface Interaction

When a light source moves, surfaces must respond with updated shading patterns that reflect both geometry and material properties.

Diffuse Shading on Surfaces. The overall shading gradients on diffuse surfaces shift according to light direction, requiring coherent large-scale brightness redistribution.

Specular Highlights. Highlights must relocate precisely across curved or planar glossy surfaces, with intensity modulated by surface normals and microfacet properties.

Material-Aware Response. Different materials like metal, plastic, paper, and fabric exhibit distinct BRDF behaviors. LightMover preserves these distinctions, producing sharp specular lobes on metallic surfaces and softer responses on matte objects.

The results in Figure 11 demonstrate that LightMover accurately captures these interactions across diverse real-world scenes.

D.2. Cast Shadows

Shadows encode critical geometric information and are highly sensitive to light movement. This makes them one of the most challenging aspects of the task.

Shadow Direction & Length. As the light source moves, cast shadows must rotate and stretch according to object–surface geometry. A physically inconsistent direction immediately breaks realism.

Shadow Softness (Penumbra / Umbra). The softness of a shadow depends on the light source size and distance.

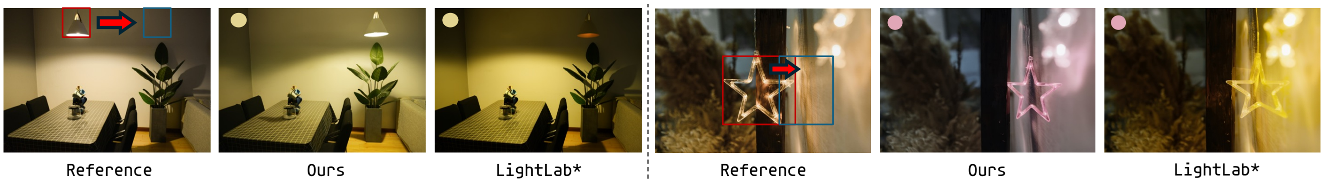
Reference Ours LightLab* Reference Ours LightLab*

Figure 3. **Comparison with the learnable embedding variant.** We evaluate the move-with-color-change task, where the colored dot in the top-left indicates the target color. The variant with learnable embeddings fails to correctly encode color conditions when movement and color changes are applied together.

{reference_image}
Move the light source from the RED box to the BLUE box, then remove all bounding boxes.

Figure 4. Concise prompt for the one-step editing.

{reference_image}
You are an expert image editor specialized in realistic light-source manipulation and relighting. Your task is to move a light source from one location to another while maintaining physical realism.
Key considerations:
1. Ensure physically accurate lighting behavior
2. Shadows and highlights must update based on the new light position
3. Preserve global scene consistency
4. Maintain light intensity and color temperature unless specified
5. Keep material properties (reflectance, specularity) unchanged

Figure 5. **Basic prompt variant for the one-step editing.** This prompt emphasizes physical correctness and globally consistent relighting.

Our model adapts shadow edges to match these conditions, producing realistic penumbra transitions.

Occlusion Reasoning. Correct shadow placement requires inferring hidden or partially visible geometry. Our method captures this structure, preserving accurate attachment points and occlusion boundaries.

As shown in Figure 12, the generated shadows are coherent, geometry-aware, and physically plausible.

D.3. Reflections

Reflections provide one of the strongest tests of illumination consistency since they depend jointly on light position, surface orientation, and scene geometry.

Reflection Direction Consistency. Reflective surfaces: wet roads, metal, polished materials must display highlights and reflected objects at positions consistent with the moved light source.

Reflection Intensity and Falloff. The brightness and attenuation of reflected highlights must adjust with the illumination energy and direction.

Environment & Object Coupling. Reflections must remain coherent with both the environment and the emitting light source, especially in urban scenes with multiple reflective cues.

{reference_image}
Reference Image Analysis:
- RED BOX: Current position of the light source
- BLUE BOX: Desired target position for the light source

Task: Light Source Relocation with Realistic Relighting
Step-by-step requirements:
1. Object Movement:
- Move the light source object from the red box to the blue box position
- Maintain the object's orientation and appearance

2. Lighting Updates:
- Calculate new shadow directions based on the target light position
- Update highlights on reflective surfaces
- Adjust ambient lighting in the scene
- Ensure shadows are cast from the new light position

3. Physical Realism:
- Maintain consistent light intensity
- Keep natural shadow softness based on light source size
- Preserve material properties of all objects
- Ensure proper light falloff with distance

4. Scene Consistency:
- Keep the background unchanged except for lighting
- Maintain color temperature
- Preserve image resolution and quality

Generate the photorealistic result with the light source successfully relocated.

Figure 6. **"Precise" step-by-step prompt variant for the one-step editing.** This version enforces explicit reasoning over shadow geometry, reflectance behavior, and scene consistency.

Figure 13 shows that LightMover maintains this global consistency, accurately updating reflective behavior under light movement.

E. Limitations

LightMover currently targets single-image editing and may produce temporally inconsistent results when applied frame-by-frame to video; video extension is a natural direction for future work. The method struggles with large-scale ambient lighting changes (e.g., relocating outdoor sunlight) due to limited training coverage, and placement of lights behind transparent or heavily occluded objects remains challenging. The approach inherits the assumptions of 2D diffusion-based editing and does not recover explicit 3D geometry or materials, which can lead to physically implausible results in scenes with complex light transport.

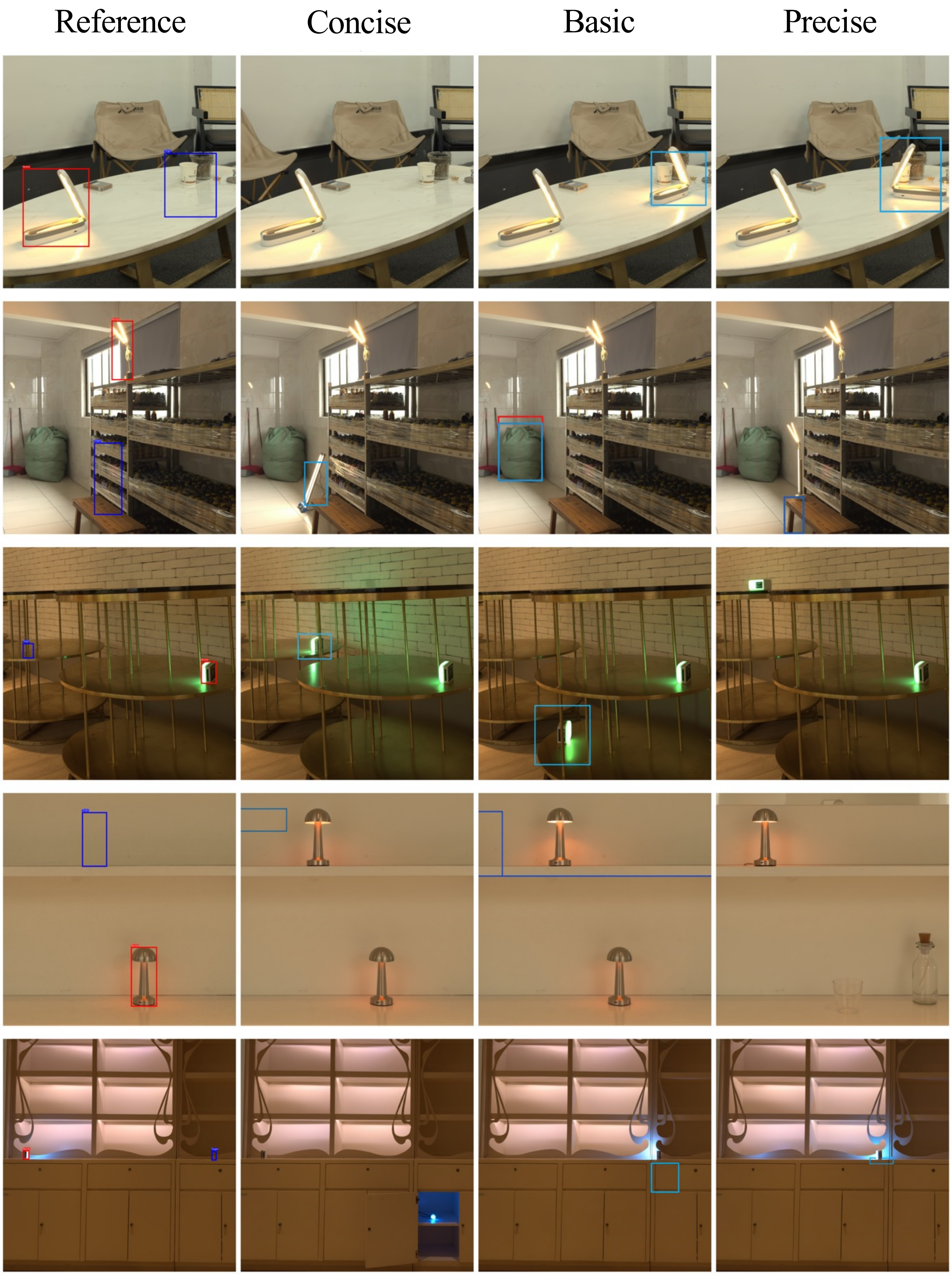

Figure 7. **Light movement results for Nano-Banana [2] one-step editing.** We show the input reference image and the results of using different prompts.

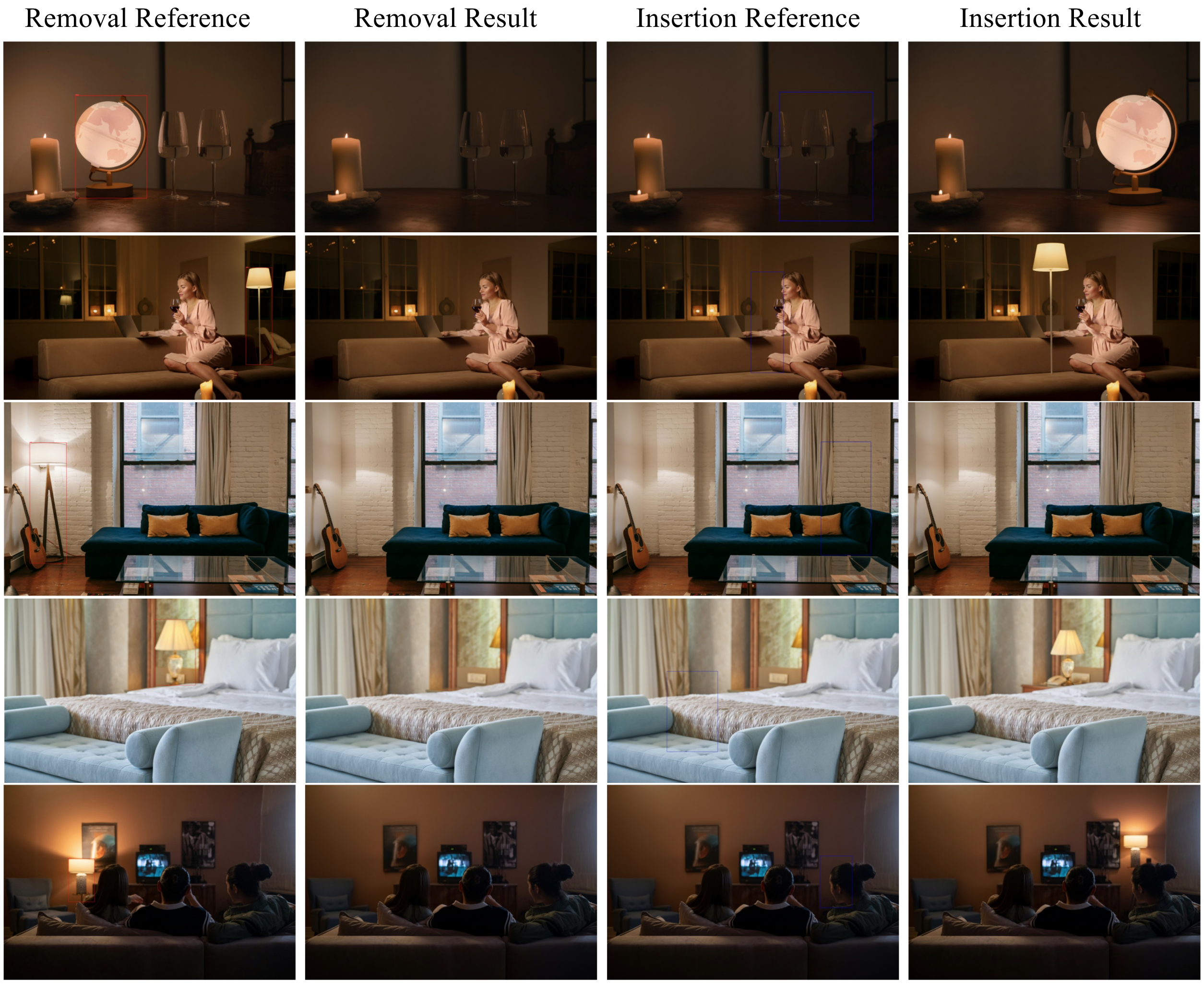

Figure 8. **Light movement results for Nano-Banana [2] two-step editing.** We show the input reference image for removal and insertion tasks, and the results.

{reference_image}
Remove the light source object located in the RED bounding box from the image.
Requirements:
1. Fill the area naturally with appropriate background content.
2. Ensure the removal is seamless and photorealistic.
3. Update shadows and lighting to reflect the absence of the light source.

Figure 9. Removal prompt for the two-step Nano-Banana editing.

References

[1] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 1

[2] Google Research. Gemini 2.5 flash — image model. `https://aistudio.google.com/models/gemini-2-5-flash-image`, 2025. AI Studio, Google Research. 4, 5

{object_image}
{reference_image}
You are provided with two images:
1. A light source object to be inserted
2. A scene image with a BLUE bounding box marking the target location

Task: Insert the light source object from the first image into the BLUE bounding box location in the second image.
Requirements:
- Place the object naturally within the blue box area
- Ensure seamless integration with the scene
- Add appropriate shadows and lighting effects from the new light source
- The result should be photorealistic with proper light-scene interactions
- Remove the blue bounding box in the final output

Figure 10. Insertion prompt for the two-step Nano-Banana editing.

Surface Interaction

(a) Diffuse Shading on Surfaces

(b) Specular Highlights

(c) Material-Aware Response

Figure 11. **Surface interaction results.** Examples of diffuse shading, specular highlights, and material-dependent responses under light movement.

Cast Shadows


(a) Occlusion Reasoning (b) Shadow Direction (c) Shadow Softness

Figure 12. **Cast shadow results.** Examples demonstrating changes in shadow direction, softness, and occlusion structure.

Reflection


Figure 13. **Reflection results.** Examples of updated highlight positions and reflective consistency after repositioning the light source.